Chemical and Biological Technologies in Agriculture

RESEARCH

Open Access

Assessment of the genetic variations of sago palm *Metroxylon sagu* in three regions of Sarawak, Malaysia using amplified fragment length polymorphism (AFLP) marker

Mehvish Nisar and Hasnain Hussain*

Abstract

Background: Sago palm (*Metroxylon sagu)* is an important staple crop in the rural area of Sarawak, Malaysia. The palms grow well in the rough, swampy peat, from low flooded areas to uplands and acidic to neutral soils. Among the starch-producing crops, this palm is the most productive and promising, where it can store a significant amount of carbohydrate in the trunk. Due to the lack of molecular study of sago palm, the work described here aimed to develop molecular markers for identifying *M. sagu* and accessing the genetic variations of sago palm in different locations of Sarawak.

Results: Amplified fragment length polymorphism was employed to determine the genetic variations of sago palm between fifteen populations collected from three different locations: Mukah, Pusa and Samarahan. Based on 35,016 polymorphic fragments generated from 32 primer combinations, polymorphic information content (PIC), marker index (MI) and resolving power (RP) averaged 0.269, 0.005157 and 51.52039 per primer combination, respectively, were obtained, signifying the effectiveness and accuracy of the AFLP marker. Mukah showed higher diversity ($I = 0.428$, $H = 0.232$), while Pusa showed the lowest ($I = 0.367$, $H = 0.240$). The average value of the Shannon information index was 0.404. AMOVA showed 99% of the variation was found within the population. The Mantel test was performed with 9999 permutations. A significantly positive correlation ($r = 0.220$, $p = 0.010$) was observed between the genetic divergence of the population (Nei genetic distance) with the geographical distance.

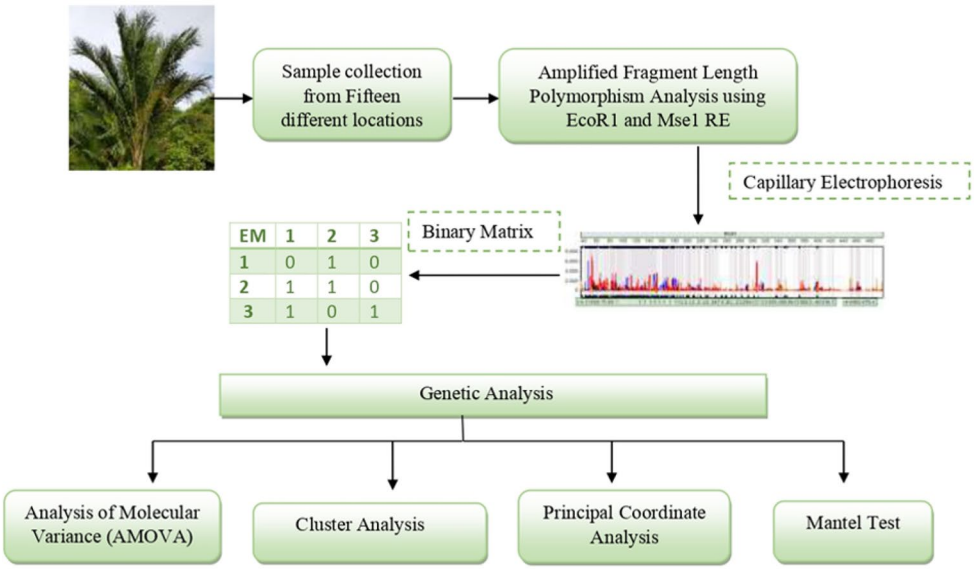
Conclusions: Our study considered AFLP is as an effective and reliable method to assess the genetic variations and the identification of *M. sagu.*

Keywords: AFLP, DNA fingerprinting, Genetic diversity, Genetic variation, *M. sagu*

*Correspondence: hhasnain@unimas.my

Centre for Sago Research (CoSAR), Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

Graphical Abstract



Introduction

Background

Sago palm (*Metroxylon sagu* Rottboel) belongs to the family Arecaceae. Among three subfamilies in the Aceraceae family, fourteen genera are known to produce starch in the trunk, and the genus *Metroxylon* is the most valuable and promising in producing starch. The distribution of the genus *Metroxylon* is from South East Asia to Melanesia, Micronesia, and Polynesia [1]. It is divided into *Metroxylon (Eumetroxylon*) and *Coelococcus* [2]. Only one species *M. sagu* Rottb. called the true sago palm, is recognized in the section Metroxylon. Sago palm grows in zone 10 north and south of the equator at the Malay Peninsula extends across Southeast Asia, in Thailand, Philippines, Malaysia, Brunei, Indonesia, and north-western Melanesia, Papua New Guinea (PNG), and the Solomon Islands regions. In Malaysia, the sago palm can be found in Sarawak, Sabah, and Johor and is cultivated mainly in Mukah and Pusa in Sarawak. Sago palm is highly adaptable to the environment and can grow from low flooded areas to uplands and acidic to neutral soils, where other major crops cannot grow well. However, sago palm will grow better in upland conditions with sufficient water and show preferable production than under submerged or flooded conditions. The wild stands of sago palm grow near lakes, on riverbanks, and in wet soil. They can be found up to an elevation of about 1000 m above sea level in PNG and about 700 m in North Sulawesi and can primarily be found growing in swamps and peat moors, which are unsuitable for any cash crop [3]. Nowadays, sago palms are found growing in lowland swamps which are not suitable for agriculture development, which does not mean that these places are the preferred habitat for sago palms [4]. Sago palm can grow well in a humid tropical environment with temperatures of 25 °C and above on average with other tolerations such as air humidity of 70% salinity and incidental light [5].

Genetic diversity is one aspect of biological diversity that is extremely important for conservation strategies [6]. The genetic variations in plants affect the higher level of biodiversity. Without genetic diversity, the population cannot adapt and survive environmental changes. Genetic variation has an impact on the higher level of biodiversity. Analysis of genetic variations between or within different species, populations, and individuals is essential for germplasm evaluation and conservation [7]. While sago palm is considered important in the region of Sarawak, it was given less attention in scientific study, especially in molecular biology. Therefore, limited information on the genetic diversity of sago palm can be found. Increasing utilities of sago palm need genetic diversity information within the species and population level. The sago palm's genetic diversity studies are essential for future breeding programs and germplasm conservation. The relationship between the genetic distance and geographical distribution of sago palms with a high level of diversity was found in the Malay Archipelago and Papua New Guinea using

the RAPD marker [8]. The genetic and morphological variations of sago palm in PNG using AFLP marker show no correlation between different morphological characteristics and genetic distances [9]. In Indonesia, molecular markers of sago palms show high variations based on RAPD and polymorphism level and genetic diversity calculation [10]. Later, Wx gene markers were used to assess the genetic variations and genetic relationship of sago palm in Indonesia to show higher variations observed in levels of individual sampling stages, followed by population and the island sampling stages [11]. Wx genes in sago palm show a high polymorphism level, with fourteen genotypes, high diversity, and high heterozygous value.

The amplified fragment length polymorphism (AFLP) technique is based on selective amplification of restriction fragments from a total digested genomic DNA to generate and compare unique fingerprints for the genome of interest. AFLP applies to all plant species and is employed in studying genetic variation within species or closely related species [12] to interfere with population-level phylogenies and biogeographic patterns to generate genetic maps [13] and determine relatedness among cultivars. AFLP is used in the genome-wide screening of genetic diversity and can be applied to almost any organism [14, 15]. Besides, AFLP is also applicable for characterizing cytosine methylation levels and target sites [16]. The rapidity and a large amount of data generated by this approach and robustness and repeatability make AFLP analysis a commonly used tool in population genetic and ecological studies [15, 15].

The AFLP markers offer several advantages over other DNA molecular techniques. The AFLP can quickly generate many fragments from any organism without prior knowledge of the genomic sequence and requires a small amount of starting template [16], hence having a relatively low start-up cost. It is highly reproducible as it combines both the reliability of RFLP and the flexibility of other PCR-based techniques. AFLP markers sometimes display dominance and are labour-intensive; despite this, they can be easily multiplexed to amplify the same batch of hundreds of genomic fragments from different individuals.

In this study, molecular markers were developed to identify *M. sagu*. The AFLP markers were used to evaluate the genetic variations of *M. sagu* among and within 15 populations collected from three different locations. The obtained information will contribute to the conservation and maintenance of such genetic resources.

Result

Polymorphic analysis

The 32 selective AFLP primer combinations produced a total of 35,016 DNA bands from 120 representative samples of 15 populations of *M. sagu*. The number of bands per primer ranged from 358 to 1592 with a polymorphic frequency of 100% with an average of 1094.28. The number of loci of 15 populations ranged from 716 to 908, with the polymorphic frequency from 66.54 to 84.48%, with an average of 78.36%. The maximum value existed in population Mukah 5 and the minimum value in population Pusa 1 (Table 1).

Marker attribute

Polymorphic information content

The powerfulness and discriminatory power of the informative AFLP profile mainly depend on three parameters; polymorphic information content (PIC), marker index (MI), and resolving power (Rp). The PIC determined for each primer was calculated, and the PIC value for the polymorphic fragments ranged from 0.096 to 0.350 with an average of 0.2678. All the primer combinations presented high discrimination power. The primer combination 56-ROXN/E-ATG/M-GTC was the most informative, with the highest PIC value of 0.350. The lowest PIC value, 0.096, was recorded in the primer combination 5HEX/E-AAT/M-GTC (Table 2).

Marker index (MI)

Marker index (MI) is a feature of marker diversity that represents the product of the effective multiplex ratio (E) and the arithmetic mean heterozygosity (Havp); therefore, the MI value for all the primer combinations was calculated. The MI value ranges from 0.000361 to 0.008072 with an average of 0.005157. The highest value was obtained with the primer combination of /56-ROXN/E-ATG/M-CAC and the lowest value for primer combination 5HEX/E-AAT/M-GTC (Table 2). A strong positive correlation was observed between MI and PIC value and was statistically significant ($r = 0.9488$, $r^2 = 0.9002$, $p = < 0.00001$).

Resolving power (RP)

Resolving power (RP) indicates the discriminatory potential of the marker combination to distinguish between large numbers of genotypes. The resolving power ranged from 17.46667 to 99.33333 with an average of 51.52039. The highest value was scored with primer combination/56-ROXN/E-ATG/M-CAA and the lowest for the combination 5HEX/E-80.

AAT/M-CTA (Table 2). The Rp value was positively correlated with MI and the values was statistically significant ($r = 0.4883$, $r^2 = 0.2384$, $p < 0.004606$).

Genetic diversity and genetic structure

The genetic diversity parameters analyzed in this study are presented in Table 3. The genetic diversity analysis

Table 1 AFLP primer sequences, number of loci, and number of polymorphic loci

Primer Code	Sequence (5′–3′)	No of loci	No of polymorphic loci	Percentage of polymorphic loci
E-ACG/M-CTT*	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CTT	891	891	100%
E-ACG/M-CAA	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CAA	839	839	100%
E-ACG/M-CAT	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CAT	1208	1208	100%
E-ACG/M-CAG	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CAG	360	360	100%
E-ACG/M-CTA	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CTA	1081	1081	100%
E-ACG/M-CAC	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CAC	1252	1252	100%
E-ACG/M-CTC	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + CTC	1269	1269	100%
E-ACG/M-GTC	/56-FAM/AGACTGCGTACCAATTC + ACG GATGAGTCCTGAGTAA + GTA	956	956	100%
E-AAT/M-CTT	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CTT	1334	1334	100%
E-AAT/M-CAA	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CAA	826	826	100%
E-AAT/M-CAT	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CAT	1359	1359	100%
E-AAT/M-CAG	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CAG	1334	1334	100%
E-AAT/M-CTA	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CTA	1305	1305	100%
E-AAT/M-CAC	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CAC	1392	1392	100%
E-AAT/M-CTA	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + CTA	1235	1235	100%
E-AAT/M-GTC	/56-ROXN/AGACTGCGTACCAATTC + AAT GATGAGTCCTGAGTAA + GTC	1592	1592	100%
E-ATG/M-CTT	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + CTT	1212	1212	100%
E-ATG/M-CAA	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + CAA	775	775	100%
E-ATG/M-CAT	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + GTC	1135	1135	100%
E-ATG/M-CAG	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + CAG	1010	1010	100%
E-ATG/M-CTA	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + CTA	1084	1084	100%
E-ATG/M-CAC	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + CAC	1184	1184	100%
E-ATG/M-CTA	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + CTA	1016	1016	100%
E-ATG/M-GTC	/55-TAMK/AGACTGCGTACCAATTC + ATG GATGAGTCCTGAGTAA + GTC	1440	1440	100%
E-AAC/M-CTT	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CTT	1407	1407	100%
E-AAC/M-CAA	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CAA	1099	1099	100%
E-AAC/M-CAT	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CAT	1209	1209	100%

Table 1 (continued)

Primer Code	Sequence (5′–3′)	No of loci	No of polymorphic loci	Percentage of polymorphic loci
E-AAC/M-CAG	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CAG	1333	1333	100%
E-AAC/M-CTA	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CTA	358	358	100%
E-AAC/M-CAC	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CAC	413	413	100%
E-AAC/M-CTC	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CTC	1016	1016	100%
E-AAC/M-GTC	5HEX/AGACTGCGTACCAATTC + AAC GATGAGTCCTGAGTAA + CTC	1092	1092	100%
Total			35,016	100%

E, *Eco*R; M, *Mse*1

showed that the overall gene diversity (Ht) was about 0.2855, the average gene diversity within the population (Hw) was about 0.3019. The average number of alleles found within each population (Na) was 1.596 (range 1.331 in the Pusa 2 population to 1.692 in Mukah 5 population. The average number of effective alleles was 1.437 (range:1.343 to 1.486). Nei's genetic diversity (Hj) of different populations ranged from 0.214 in Pusa 6 population to 0.292 in Mukah 5 and Samarahan 3 population with an average of 0.266, and the average Shannon's information index (I) was 0.404 ranged from 0.330 in the Pusa population 1 to 0.442 in Mukah population 5. All populations showed a mean of 0.186 for the heterozygosity ranging from 0.142 in Pusa 1 population to 0.209 population of Mukah 5.

Non-hierarchical analysis of genetic variability was performed using molecular variance (AMOVA) analysis to know the variation among and within different populations from three regions. The analysis indicated that the variation within populations accounted for 99% of the total variation, while among regions, variations contributed for only 1%. There were no variations among populations (Table 4). The 1% variation among regions was due to Pusa populations. There were no variations between Mukah and Samarahan. The AMOVA results indicate that the differentiation mainly existed within the population. The source of variation among the region's PhiRT value was significantly different ($p \leq 0.001$). The source of variation among populations within a region (PhiPR) and within individuals (PhiPT) was significantly indifferent. Fst calculated by AFLP–SURV as an indication of population differentiation was significantly indifferent ($p > 0.05$). The total Fst value was -0.0491, which indicates no genetic difference between the populations of *M. sagu* in Sarawak, except a minimal variation in sago palms of the Pusa region (Table 4).

The pairwise population of Nei's genetic identity and distance shows a high genetic identity of 0.993 and a low genetic distance of 0.007 among all populations.

Cluster analysis

Genotyping data obtained from all the polymorphic fragments from 32 primer combinations were used to estimate pairwise similarity comparisons among the fifteen populations. A distant genetic matrix was calculated using Nei genetic distant coefficient using the GenAlex program. Therefore, the Nei genetic distant matrix was used to construct the unweighted pair group method arithmetic (UPGMA) dendrogram visualized through the Xlstat program. Nei's genetic distance coefficient value between the populations was in the range of 0.006–0.014, which showed a very low genetic divergence among different populations of *M. sagu*. The genetic distance-based UPGMA dendrogram (Fig. 1) showed two distinctly major clusters at a distant coefficient of 0.010. Clusters I comprised four populations from Pusa locations. Cluster II contains all other populations from all three regions. As a whole, for clustering results, most of the populations cluster together, representing a closer relationship between all the three populations. Although four among five populations from Pusa cluster together in cluster I.

A genetic distant matrix obtained based on Nei's distant coefficient was also subjected for principal coordinate analysis (PCoA). As shown in Table 5 and Fig. 2, the plot of the first principal axes represents 45.53% of the variation. The second and the third principal axes represent 11.74% and 9.09%, respectively, giving a cumulative variation of 66.36%. Populations in diverse groups were intermixed, and PCoA failed to resolve accurately according to their geographic origin.

Table 2 Marker parameters calculated for each AFLP primer

S.no	Markers	PIC	MI	Dp	Rp
1	/56-FAM/E-ACG/M-CTT	0.269775	0.004312	0.959557	63.86667
2	/56-FAM/E-ACG/M-CAA	0.251065	0.003521	0.967859	80.8
3	/56-FAM/E-ACG/M-CAT	0.299239	0.005898	0.941728	73.6
4	/56-FAM/E-ACG/M-CAG	0.146313	0.000922	0.992441	23.73333
5	/56-FAM/E-ACG/M-CTA	0.286748	0.346673	0.212367	38.13333
6	/56-FAM/E-ACG/M-CAC	0.315441	0.007015	0.928167	74.26667
7	/56-FAM/E-ACG/M-CTC	0.322966	0.007614	0.920516	70
8	E-/56-FAM/ACG/M-GTC	0.302646	0.006115	0.939153	39.86667
9	5HEX/E-AAT/M-CTT	0.296895	0.005753	0.94343	53.73333
10	5HEX/E-AAT/M-CAA	0.300536	0.00598	0.940767	67.6
11	5HEX/E-AAT/M-CAT	0.27136	0.004385	0.958768	51.6
12	5HEX/E-AAT/M-CAG	0.290206	0.005359	0.947973	36
13	5HEX/E-AAT/M-CTA	0.107383	0.00046	0.996342	17.46667
14	5HEX/E-AAT/M-CAC	0.109021	0.000476	0.996213	17.73333
15	5HEX/E-AAT/M-CTA	0.109021	0.000476	0.996213	17.73333
16	5HEX/E-AAT/M-GTC	0.096133	0.000361	0.997152	22.13333
17	56-ROXN/E-ATG/M-CTT	0.326912	0.007953	0.916067	50.8
18	56-ROXN/E-ATG/M-CAA	0.302806	0.006126	0.93903	99.33333
19	56-ROXN/E-ATG/M-CAT	0.31406	0.006911	0.929467	58.53333
20	56-ROXN/E-ATG/M-CAG	0.308657	0.006521	0.934272	41.2
21	56-ROXN/E-ATG/M-CTA	0.315472	0.007017	0.928139	37.46667
22	56-ROXN/E-ATG/M-CAC	0.328254	0.008072	0.914473	45.86667
23	56-ROXN/E-ATG/M-CTA	0.282019	0.004913	0.95299	57.46667
24	56-ROXN/E-ATG/M-GTC	0.350166	0.01044	0.880327	57.06667
25	/55-TAMK/E-AAC/M-CTT	0.279974	0.004807	0.954159	72.53333
26	/55-TAMK/E-AAC/M-CAA	0.211887	0.00225	0.980409	43.46667
27	/55-TAMK/E-AAC/M-CAT	0.336602	0.008873	0.903511	68
28	/55-TAMK/E-AAC/M-CAG	0.267617	0.004213	0.960614	30.66667
29	/55-TAMK/E-AAC/M-CTA	0.297591	0.005795	0.942928	38.13333
30	/55-TAMK/E-AAC/M-CAC	0.305474	0.006303	0.936912	52.26667
31	/55-TAMK/E-AAC/M-CTC	0.244021	0.00349	0.970543	78.28571
32	/55-TAMK/E-AAC/M-GTC	0.323617	0.007668	0.919805	38.93333
	Average	0.267809	0.005157	0.948734	51.52039

Table 3 Genetic parameters of 15 populations of *M. sagu*

Populations	Na	Ne	*I*	Hj	He
Mukah1	1.635	1.462	0.423	0.279	0.197
Mukah2	1.592	1.432	0.405	0.265	0.186
Mukah3	1.681	1.480	0.438	0.289	0.204
Mukah4	1.662	1.467	0.430	0.283	0.201
Mukah5	1.692	1.486	0.442	0.292	0.209
Pusa1	1.331	1.343	0.330	0.214	0.142
Pusa2	1.558	1.439	0.403	0.266	0.186
Pusa3	1.376	1.366	0.347	0.226	0.154
Pusa4	1.513	1.410	0.385	0.252	0.174
Pusa5	1.457	1.390	0.368	0.240	0.166
Samarahan1	1.511	1.419	0.388	0.255	0.177
Samarahan2	1.686	1.485	0.441	0.292	0.208
Samarahan3	1.543	1.431	0.398	0.262	0.184
Samarahan4	1.647	1.467	0.427	0.282	0.196
Samarahan5	1.647	1.478	0.433	0.286	0.202
Total	1.569	1.437	0.404	0.266	0.186

Na, observed number of alleles; Ne, effective number of alleles; Hj, Nei genetic diversity; I, Shannons index; He, expected heterozygosity

The Mantel test (Fig. 3) was performed with 9999 permutations was performed. A significantly positive correlation ($r=0.220$, $p=0.010$) was observed between the genetic divergence of the population (Nei genetic distance) with the geographical distance.

Discussion

High-throughput AFLP markers combined with a DNA-bulking approach were used to develop the molecular marker and evaluate the genetic variation of *M. sagu*. The AFLP technique is a highly reproducible method, which combines the reliability of RFLP and the power and sensitivity of the PCR method [18] to produce highly polymorphic markers. AFLP has been successfully used to investigate the genetic diversity and phylogenetic relationship between different species and closely related genotypes [12].

In the case of *M. sagu*, relatively less attention has been given to it. Therefore, a few studies have been conducted to assess molecular diversity using different markers. Sago palm shows genetic variations using RAPD markers in the progenies derived from natural pollinations [19].

Table 4 Summary of analysis of molecular variance (AMOVA) based on AFLP marker

Source	d.f	SS	MS	Est. Var	%	Phi	*p* value
Among Regions	2	278.800	139.400	1.657	1%	PhiRT = 0.011	0.034
Among Pops	12	877.550	73.129	0.000	0%	PhiPR = − 0.074	1.000
Within Pops	105	17,152.875	163.361	163.361	99%	PhiPT = − 0.063	1.000
Total	119	18,309.225		165.017	100%		

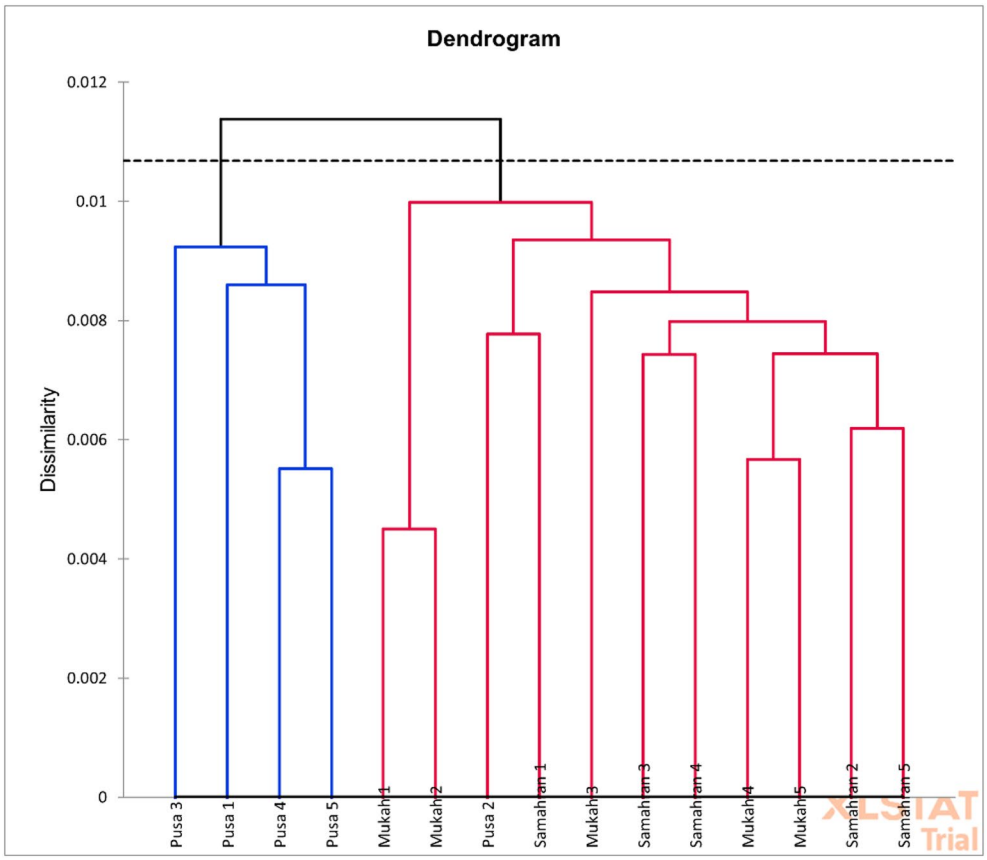


Fig. 1 Dendrogram generated using unweighted pair group method with arithmetic average (UPGMA) analysis shows a relationship between 15 *M. sagu* population using a database on Nei's genetic distance

Table 5 Principal coordinate analysis (PCoA) of 15 M. *sagu* populations on three axis

Principal coordinate	Eigenvalue	Variability	Cumulative variance
F1	0.0009624	45.531	45.531
F2	0.00024824	11.744	57.275
F3	0.0001923	9.0975	66.3725

AFLP profiling and marker polymorphism

The efficiency of any molecular marker technique depends on the amount of polymorphism detected [20]. In the present study, high-quality marker profiling was obtained using 32 AFLP *Eco*R1 and *Mse*1 primers combinations on 15 *M. sagu* samples collected from different locations. These AFLP markers generated 100% polymorphism, which was confirmed by a DNA profiling and genetic diversity analysis. All the primer combinations used were found polymorphic and generated 35,016 differently sized alleles with an average of 1094.25 per primer across all the populations of *M. sagu*. This study of *M. sagu* using AFLP markers showed more number of alleles as compared to the previous studies of *M. sagu* using RAPD

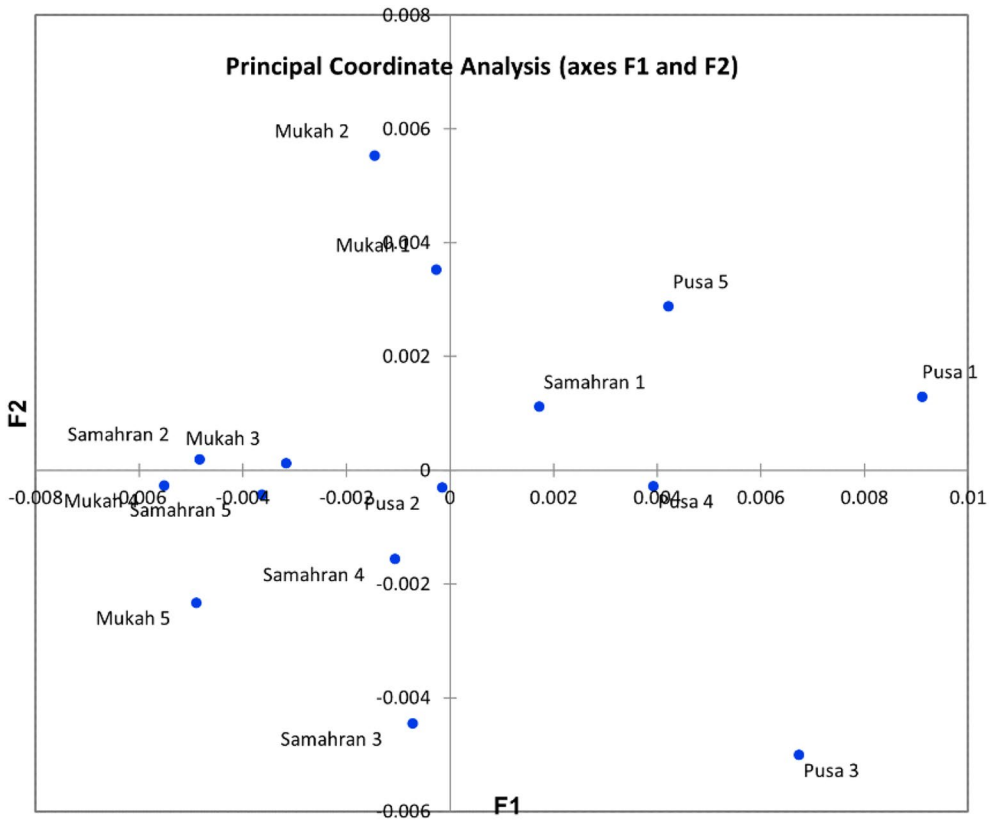


Fig. 2 Principal coordinate analysis (PCoA) of the 15 M. *sagu* populations based on AFLP markers

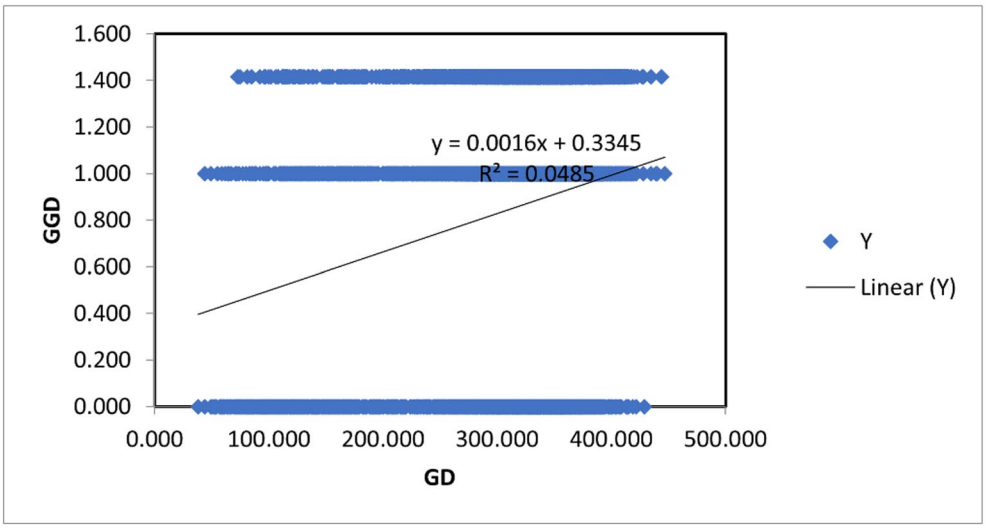


Fig.3 Mantel test correlating geographic distance and genetic distance

(Random amplified polymorphic DNA), in which 83 number polymorphic fragments were reported [10], and the average polymorphism of 93% was reported [8], indicating that AFLP markers are useful in the genetic variation study of *M. sagu.* The number of loci of 15 populations ranged from 716 to 908, with

the polymorphic frequency from 66.54 to 84.48%. The maximum value existed in population Mukah and the minimum value in Pusa.

All these results indicated that high polymorphism existed in the population of *M. sagu*. The percentage of polymorphic markers become reliable when a large number of loci are generated [21, 22]. The most significant advantage of this technique is to generate a large number of polymorphic DNA fragments, as the accuracy of measurements of genetic distance increases with the number of loci used [23]. In a study, twenty *Eco*R1 + *Mse*1 primer combinations were used in 13 Argan tree populations, which showed 100% polymorphism [24]. The genetic diversity of 127 individuals from endangered species of *Glehnia littoralis* using ten pairs of AFLP primers generated a total of 1929 bands with 100% polymorphism [25]. In comparison, the present study detected a high polymorphism and provided higher resolution loci. High-resolution loci were generated using an electrophoresis system with different fluorescent-labelled primers and laser detectors in ABI genetic analyzer, which have higher detection power than the ordinary PAGE system.

Discriminatory power of AFLP primer combination

The discriminatory power of AFLP markers to assess the genetic diversity has been identified by several parameters, such as PIC, RP, and MI [26, 27]. However, the PIC value has been used predominantly to determine the discriminatory power or informativeness of markers in a majority of diversity studies [28]. The concept of Rp (resolving power) was proposed to evaluate the discriminatory power of the AFLP primer combination [29]. MI (Marker Index) is used to calculate the efficiency of markers [30].

The PIC value of different fragments obtained by different primer combinations was 0.096–0.350 (average = 0.269). The maximum value of PIC for the dominant marker such as AFLP can be expected below 0.50 [20]. The PIC value, which we found in our study, is considered informative. MI together with PIC value has been used to assess the informativeness of AFLP primer combinations in various plant species, e.g., [31], Orchardgrass (PIC = 0.252, MI = 16.32, *Elymus tangutorum* (0.250, MI = 23.07) [32]. Our study found that the MI values were in the range of 0.000361 to 0.008072 (average = 0.005157). Resolving power was used to assess the discriminatory power of different AFLP combinations. The primer combination used in our study showed the RP value in the range of 17.46–99.33 (average = 51.52039).

These indexes have been used to compare AFLPs with other molecular markers. A strong linear relationship has been found between the capability of primer to discriminate genotypes and the resolving power (Rp), but not the marker index (MI) [33, 34]. In our study a strong positive correlation was found between PIC and MI ($r = 0.9488$, $r^2 = 0.9002$, $p = < 0.00001$), also there was a positive correlation between MI and RP ($r = 0.4883$, $r^2 = 0.2384$, $p < 0.004606$).In a study, it was reported that a strong correlation was found between MI and RP ($r^2 = 0.99$, $p < 0.005$), which signifies that the discriminatory power of primer combinations could also be determined by [35, 36]. These results indicate that the AFLP markers are highly reliable, discriminative, and powerful markers. Markers with ROX fluorescent dye have proven to be best to determine the informativeness compared to other combinations.

Genetic diversity and genetic structure

The Shannon index is a precise alternative measure of diversity due to no need to estimate allele frequencies under Hardy–Weinberg equilibrium. The Shannon index may vary from 0 to 1, and lower genetic diversity is represented by a value closer to zero [37]. In this study, the mean Shannon's Index (I) (0.404) was higher in other plants using different molecular markers. *Elymus sibiricus* germplasm from Western China showed $I = 0.2850$ using SCoT marker [38] and $I = 0.237$ using EST–SSR marker. The genetic diversity and the population structure are revealed by environmental factors, such as geographic distribution, life cycle, selection, and adaption [39]. Considering the genetic differences among studied regions, the populations from Mukah showed higher diversity ($I = 0.428$, $H = 0.282$), while the Pusa population showed the lowest ($I = 0.367$, $H = 0.240$).

The genetic diversity of the population is a foremost ability to adapt to a variety of changing environmental conditions [40]. However, in this study, AMOVA analysis showed that the major proportion (99%) of the total genetic variation of *M. sagu* was found within a population rather than among populations and geographical regions. Even though it accounted for only 1% of the variation, the variation between regions was statistically significant. In the previous study on sago palm using cpDNA markers, the highest percentage of the population (95%) was also observed within a population rather than among population and island [41]. AMOVA analysis also showed 2% variation among regions between Pusa and Mukah and 1% variation among regions between Pusa and Samarahan, while there is no variation among regions between Mukah and Samarahan. The PhipT / FsT value is very low (0.), equal to zero. This indicates no genetic difference between the population of *M. sagu* in Sarawak. This result is also supported by the low Nei's genetic distance (0.007) between all the fifteen populations and high Nei's genetic similarity (0.993).

The cluster analysis using the UPGMA approach grouped fifteen *M. sagu* populations into three groups. Cluster I grouped three populations from all the three locations, pop 7 from Pusa and population 11 from Samarahan together in one subgroup and population 2 from Mukah separately in another subgroup. Cluster II grouped eight populations from two locations, and four populations were from Mukah and four from Samarahan. Cluster III grouped all the four populations from Pusa together. The UPGMA dendrogram based on Nei's genetic distance did not show a clear separation of the 15 populations of sago palm from three different locations. This result was supported by a study by [1] in which the *M. sagu* species cluster into one group. The PCoA, which explained 66.36% of total variation on three components, also supports the UPGMA analysis, which group the different populations together. This also indicates no genetic difference between the different populations of *M. sagu* in Sarawak. In a study of *M. sagu* in Papua, 14% of the total variation was accounted on two principal components and PCoA, which indicated that the accession from the same locality showed a weak tendency to group together [9]. The stepwise Mantel test indicates a positive correlation between genetic and the geographic distance of the *M. sagu* in three regions of Sarawak [8, 9] also showed a strong connection between genetic distance and geographic distance. According to AMOVA and UPGMA analysis results, the Pusa population slightly differs from Mukah and Samarahan, and the Pusa population is closer to Samarahan than Mukah. This study shows that AFLP markers can be considered a powerful and reliable method to assess the genetic variation and identify the *M. sagu*. The high level of diversity within population and the low level of genetic diversity among the populations occurred probably by the limited sample size. These observations showed that the site of within population should be the main focus for conservation and sustainable rather than the site among population and regions.

Conclusions

This study aimed to develop molecular markers and study the genetic variation of *M. sagu* from different locations of Sarawak (Mukah, Pusa and Samarahan). Prior to the marker study, evaluation of the DNA extraction method showed that the CTAB method is best for the isolation of pure DNA for the *M. sagu*. These AFLP markers generated 100% polymorphism, which was confirmed by a DNA profiling and genetic diversity analysis. All the statistical analyses, including AMOVA, mental tests, UPGMA, PCoA, showed genetic diversity and suggested that the sago palm in Sarawak shows no difference genetically. They belongs to the same *M. sagu* species. This study showed that amplified fragment length polymorphism (AFLP) is considered the most powerful and suitable molecular marker for studying the genetic variations of *M. sagu.*

Methods

Sample collection

In this study, the plant material chosen was the leaf sample from sago palm (*M. sagu).* Leaf samples were collected from five different localities in each Pusa, Mukah and Samarahan. GPS reference WGS 84 datum for each palm sampled and the phenotypic features of

Table 6 Sample location based on places and GPS reference WGS 84

	Sample Name / populations	Location	Location	Elevation	Phenotypic features leaf
1	Mukah 1	Dalat	1°27′19.5"N 111°33′56.8"E	39 m	Long and Spineless
2	Mukah 2	Dalat/Oya	2°51′25.3"N 111°49′54.8"E	14 m	Long and Spineless
4	Mukah 4	Balingian	2°59′44.3"N 112°29′08.6"E	17 m	Long With spines
5	Mukah 5	Mukah Town	2°53′31.9"N 112°04′18.1"E	39 m	Long Spineless
6	Pusa 1	Tambak Pusa	1°36.521"N 111°19′144"E	5 m	Long With spines
7	Pusa 2	Tambak Pusa	1°36′537"N 111°19′146"E	5 m	Long With spines
8	Pusa 3	Tambak Pusa	1°36′520"N 111°19.155"E	5 m	Long Spineless
9	Pusa 4	Tambak Pusa	1°36′51.4"N 111°19′157"E	5 m	Long Spineless
10	Pusa 5	Tambak Pusa	1°36′30.8"N 111°19′09.4"E	5 m	Long Spineless
11	Samarahan 1	Kampung Pinang	1°25′29.7"N 110°26′31.6"E	2 m	Long Spineless
12	Samarahan 2	Kampung Pinang	1°25′13.1"N 110°26′54.4"E	3 m	Long With spines
13	Samarahan 3	Asajaya	1°25′35.7"N 110°25′55.4"E	2 m	Long Spineless
14	Samarahan 4	Kampung Melaban	1°26′18.1"N 110°26′16.5"E	4 m	Long With spines
15	Samarahan 5	Kampung Meranek	1°26′15.4"N 110°27′35.4"E	6 m	Long Spineless

Table 7 Selective PCR amplification primer combinations, *Eco*RI are fluorescent-labeled

	*Eco*RI(P1+3)	*Eco*RI (P2+3)	*Eco*RI (P3+3)	*Eco*RI (P4+3)
	ACG	AAT	ATG	AAC
*Mse*I(P1+3)	/56-FAM/ P1+P1	/56-ROXN/ P1+P2	/55-TAMK/ P1+P3	/5-HEX/ P1+P4
CTT	CTT+ACG	CTT+ ATT	CCT+ATG	CCT+AAC
*Mse*I(P2+3)	P2+P1	P2+P2	P2+P3	P2+P4
CAA	CCA+ACG	CCA+ATT	CCA+ ATG	CCA+AAC
*Mse*I(P3+3)	P3+P1	P3+P2	P3+P3	P3+P4
CAA	CCA+ACG	CCA+ATT	CCA+ATG	CCA+AAC
*Mse*I(P4+3)	P4+P1	P4+P2	P4+P3	P4+P4
CAG	CAG+ACG	CAG+ATT	CAG+ATG	CAG+AAC
*Mse*I(P5+3)	P5+P1	P5+P2	P5+P3	P5+P4
CTA	CTA+ACG	CTA+ATT	CTA+ATG	CTA+AAC
*Mse*I(P6+3)	P6+P1	P6+P2	P6+P3	P6+P4
CAC	CAC+ACG	CAC+ATT	CAC+ATG	CAC+AAC
*Mse*I(P7+3)	P7+P1	P7+P2	P7+P3	P7+P4
CTC	CTC+ACG	CTC+ATT	CTC+ATG	CTC+AAC
*Mse*I(P8+3)	P8+P1	P8+P2	P8+P3	P8+P4
GTG	GTG+ACG	GTG+ATT	GTG+ATG	GTG+AAC

leaf samples is mentioned in Table 6. The sago plantation from Mukah, Pusa and Samarahan is shown in Fig. 4. All the samples were taken in the Rosette stage of the sago palm. The leaves were collected from the third latest developed frond of the sago plant, wiped with 70% ethanol, and kept immersed in liquid nitrogen in a 50 mL centrifuge tube. The samples were removed from liquid nitrogen in the lab and kept at −80 °C until needed for processing.

Sample preparation

The sago leaves were taken out from a −80 °C freezer, washed with distilled water, and wiped with 70% ethanol. The midrib of the leaves was removed, weighed, and cut into small pieces using sterile scissors. 1.6 g of leaf sample was grounded into a fine powder using motor and pestle in the presence of liquid nitrogen. The powder was transferred into a 50 mL polypropylene Falcon tube and put on ice until used from DNA isolation.

DNA extraction

The Genomic DNA was extracted and isolated using the CTAB extraction method with some modifications. A total of 3 replicates for each sample were carried out prior to genomic DNA extraction. Recovered DNA was checked for purity and integrity using a bio-photometer and standard 1% gel electrophoresis with ethidium bromide staining. The DNA concentration was adjusted to 500 ng/5.5 µl volume for polymerase chain reactions (PCRs).

AFLP analysis

AFLP fingerprinting was performed following the method described by Vos et al., 1995 [42] with some modifications. The primers were labelled with fluorescent dyes at the 5'end. 500 ng of genomic DNA was double digested with the restriction enzymes *Eco*R1 and *Mse*1(NEB), following the ligation of *Eco*R1 and *Mse*1 adapters using T4 Ligase (NEB) to generate template DNA for PCR reactions. After successful restriction ligation reaction, pre-selective amplification was carried out with *Eco*R1 + 1 (GACTGCGTACCAATTC + A) and *Mse*1 + 1 (GATGAGTCCTGAGTAA + C) primers each having one selective nucleotide. The pre-selective amplified products were used as the selective PCR amplification reaction template. Selective amplification was conducted using the four *Eco*R1 primers, each having three different extra nucleotide bases and eight *Mse*1 primers, each with three different extra nucleotide bases. *Ecor*1 primers were fluorescently labelled at 5' end with different fluorophores. A total of 81 primer combinations were tested on the DNA of each 15 individuals of *M. sagu*, and 32 primer combinations per sample were chosen for the selective PCR amplification to detect the polymorphism among the fifteen populations of *M. sagu* (Table 7).

Capillary electrophoresis of selective amplification products was performed using ABI genetic analyzer. The amplified AFLP fragment profile was analyzed with the GeneMarker 3.0.1 analysis software (Softgenetics, USA). The amplified fragment profiles were assembled in binary format and visually scored for allele presence (1) and absence (0) for all the samples. The threshold for allele calling was set at 100 relative fluorescence (rfu) so that any peaks at 100 (rfu) or higher were assigned as 1, and those lower than 1 were assigned as 0.

Fig. 4 Distribution of the field sampling in Sarawak

Data analysis

The discriminatory power of each AFLP primer combination was evaluated by calculating polymorphic information content (PIC), marker index (MI), and resolving power (RP) with polymorphic bands only using the online IMAC software. Several genetic diversity related parameters, Hj (Nei's gene diversity), Ht (the total gene diversity) and Hw (the average gene diversity) within and among populations were calculated using AFLP–SURV v.1.0 and GenALEx 6.5. The total number of bands, number of polymorphic loci, percentage of polymorphic loci (%), observed (Na) and effective number of alleles (Ne), Shannon's information index (I) and AMOVA (analysis of molecular variance) were calculated using the GenALEx 6.5. Cluster analysis of population was conducted to construct the dendrogram using UPGMA (unweighted pair group method with arithmetic mean) based on Nei's genetic distance visualized through the Xlstat program. Principal coordinate analysis (PCoA) of all the population was conducted using GenALEx 6.5. Mental test with 9999 permutations was observed between the genetic and the geographical distances using GenALEx 6.5.

Abbreviations

AMOVA: Analysis of molecular variance; ISSR: Inter-simple sequence repeats; PIC: Polymorphism information content; RAPD: Random amplified polymorphic DNA; UPGMA: Unweighted Pair Group with Arithmetic Average; AFLP: Amplified fragment length polymorphism; MI: Marker index; RP: Resolving power; PCoA: Principal coordinate analysis.

Acknowledgements

This work was supported by the Ministry of Higher Education Malaysia FRGS Grant (FRGS/1/2017/STG04/UNIMAS/02/1) and UNIMAS Tun Openg Sago Chair Research Grant (F07/TOC/1743/2018). MN is the recipient of Universiti Malaysia Sarawak Zamalah scholarship.

Author contributions

MN planned and performed the experiments, data analysis, and the main author of the manuscript. HH designed the overall study and a major contributor in manuscript preparation. All authors have read and approved the manuscript.

Funding

Ministry of Higher Education Malaysia FRGS Grant (FRGS/1/2017/STG04/UNIMAS/02/1) and UNIMAS Tun Openg Sago Chair Research Grant (F07/TOC/1743/2018) funded this research. Open Access funding provided by Universiti Malaysia Sarawak.

Availability of data and materials

All data and materials are presented in this manuscript.

Declarations

Ethics approval and consent to participate
Plant materials were obtained from commercial sago palm farms in Mukah, Pusa and Samarahan. Permission was obtained from the plantation company and landowners. The collection of plant in the present study complies with institutional guidelines.

Consent for publication
Not applicable.

Competing interests
All authors declare no interest. All authors declare no competing interests as defined by SpringerNature, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Received: 9 February 2022 Accepted: 23 June 2022
Published online: 13 July 2022

References

1. Ehara H, Toyoda Y, Johnson Dennis V. Sago Palm: multiple contributions to food security and sustainable livelihoods. Cham: Springer; 2018.
2. Rauwerdink JB. An essay on Metroxylon, the Sago Palm. Principes J Int Palm Society. 1986;30(4):165–80.
3. Ehara H. Genetic Variation and Agronomic Features of Metroxylon Palms in Asia and Pacific. In: Ehara H, Toyoda Y, Johnson DV, editors. Sago Palm: Multiple Contributions to Food Security and Sustainable Livelihoods. Singapore: Springer Singapore; 2018. p. 45–59.
4. Ehara H. *The sago palm : the food and environmental challenges of the 21st Century . The Society of Sago Palm Studies*, ed. Society of Sago Palm. Kyoto University Press, Kyoto: Trans Pacific Press, Balwyn North, Victoria. 2015.
5. Flach M. Sago palm: *Metroxylon sagu* Rottb.-Promoting the conservation and use of underutilized and neglected crops. 13. Bioversity International. 1997.
6. Kaljund K, Jaaska V. No loss of genetic diversity in small and isolated populations of *Medicago sativa* subsp. falcata. Biochem System Ecol. 2010;38:510–20.
7. Duran C, Appleby N, Edwards D, Batley J. Molecular genetic markers: Discovery, applications, data storage and visualisation. Current Bioinform. 2009;4:16–27. https://doi.org/10.2174/157489309787158198.
8. Ehara H. Relationship between geographical distribution and genetic distance of Sago Palms in the Malay Archipelago. Sago Palm. 2003;11:8–13.
9. Kjaer A, et al. Investigation of genetic and morphological variation in the Sago Palm (Metroxylon sagu; Arecaceae) in Papua New Guinea. Ann Bot. 2004;94:109–17.
10. Abbas B, et al. Genetic relationship of Sago Palm (Metroxylon sagu Rottb.) in Indonesia based on RAPD markers. Biodiversitas. 2009;10:1412–33.
11. Abbas B. Assessment genetic variation and relationship of sago palm (Metroxylon sagu Rottb.) in Indonesia based on specific expression gene (Wx genes) markers. Afr J Plant Sci. 2012. https://doi.org/10.5897/AJPS12.015.
12. Althoff DM, Gitzendanner MA, Segraves KA. The utility of amplified fragment length polymorphisms in phylogenetics: a comparison of homology within and between genomes. Syst Biol. 2007;56(3):477–84.
13. Srinivasan J, et al. An integrated physical and genetic map of the nematode *Pristionchus pacificus*. Mol Genet Genomics. 2003;269(5):715–22.
14. Bensch S, Åkesson M. Ten years of AFLP in ecology and evolution: why so few animals? Mol Ecol. 2005;14:2899–914.
15. Meudt HM, Clarke AC. Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends Plant Sci. 2007;12(3):106–17.
16. Vuylsteke M, Peleman JD, van Eijk MJ. AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat Protoc. 2007;2(6):1399–413.
17. Mueller UG, Wolfenbarger LL. AFLP genotyping and fingerprinting. Trends Ecol Evol. 1999;14(10):389–94.
18. Hussain H, Nisar M. Assessment of plant genetic variations using molecular markers: A review. J Appl Biol Biotech. 2020;8(05):099-109. https://doi.org/10.7324/JABB.2020.80514.
19. Abbas B, Dailami M, Listyorini F. Genetic variations and relationships of Papua's Endemic Orchids based on RAPD markers. Nat Sci. 2017;09:377–85.
20. Tatikonda L, et al. AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant. Plant Sci. 2009;176(4):505–13.
21. Mohammadi S, Prasanna B. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci. 2003;43:1235.
22. Assaeed AM, et al. Evaluation of genetic diversity of *Panicum turgidum* Forssk from Saudi Arabia. Saudi J Biol Sci. 2018;25(1):123–9.
23. Travis SE, Maschinski J, Keim P. An analysis of genetic variation in *Astragalus cremnophylax* var. cremnophylax, a critically endangered plant, using AFLP markers. Mol Ecol. 1996;5(6):735–45.
24. Pakhrou O, et al. Study of genetic diversity and differentiation of argan tree population (*Argania spinosa* L.) using AFLP markers. Aus J Crop Sci. 2016;10:7680.
25. Li B, et al. Genetic diversity and population structure of endangered *Glehnia littoralis* (Apiaceae) in China based on AFLP analysis. Biotechnol Biotechnol Equip. 2019;33(1):331–7.
26. Barakat M, et al. Assessment of genetic diversity among wheat doubled haploid plants using TRAP markers and morpho-agronomic traits. Aust J Crop Sci. 2013;7:104–11.
27. Quintero V, et al. Genetic structure of *Jatropha curcas* L. in Mexico and probable centre of origin. Biomass Bioenergy. 2013;60:147.
28. Varshney R, et al. Genic molecular markers in plants: development and applications. In: Varshney RK, Tuberosa R, editors., et al., Genomics-assisted crop improvement. Dordrecht: Springer, Netherlands; 2007.
29. Gupta SK, Souframanien J, Gopalakrishna T. Construction of a genetic linkage map of black gram, *Vigna mungo* (L.) Hepper, based on molecular markers and comparative studies. Genome. 2008;51(8):628–37.
30. Ghasemi Ghehsareh M, et al. Application of ISSR markers to analyze molecular relationships in Iranian Jasmine (Jasminum spp.) Accessions. Molecular Biotechnol. 2014;57:65.
31. Jiao SQ, et al. Development of novel EST-SSR markers for Ephedra sinica (Ephedraceae) by transcriptome database mining. Appl Plant Sci. 2019;7(1): e01212.
32. Wu WD, et al. Genetic diversity and structure of Elymus tangutorum accessions from western China as unraveled by AFLP markers. Hereditas. 2019;156:8.
33. Wang IJ, Bradburd GS. Isolation by environment. Mol Ecol. 2014;23(23):5649–62.
34. Prevost A, Wilkinson MJ. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor Appl Genet. 1999;98(1):107–12.
35. Gupta A, et al. AFLP based genetic relationship and population structure analysis of Canna—An ornamental plant. Sci Hortic. 2013;154:1–7.
36. Wati C, et al. Diversity analysis in Indian genotypes of linseed (*Linum usitatissimum* L.) using AFLP markers. Gene. 2014;549:171.
37. da Silva L, et al. Assessing the genetic diversity and population structure of the endangered *Chascolytrum bulbosum* (Poaceae, Poeae) using AFLP markers. Biochem Syst Ecol. 2016;68:236–42.
38. Xie W-G, et al. Diversity comparison and phylogenetic relationships of cocksfoot (*Dactylis glomerata* L.) germplasm as revealed by SSR markers. Can J Plant Sci. 2010;90(1):13–21.
39. Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol. 2004;13:1143–55.
40. Dawson TP, et al. Beyond predictions: biodiversity conservation in a changing climate. Science. 2011;332(6025):53–8.
41. Abbas B, et al. Genetic diversity of sago palm in Indonesia based on chloroplast DNA (cpDNA) markers. Biodiversitas. 2010. https://doi.org/10.13057/biodiv/d110302.
42. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 1995;23(21):4407–14. https://doi.org/10.1093/nar/23.21.4407.

Publisher's Note